DOCUMENT RESUME

ED 334 249 TM 016 821

AUTHOR Hall, Janie L.; Kleine, Paul F.
TITLE Preparing Students To Take Standardized Tests: Have We Gone Too Far?
PUB DATE 90
NOTE 17p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Elementary School Teachers; Elementary Secondary Education; *Instructional Materials; Mail Surveys; *Principals; Questionnaires; School Policy; School Surveys; *Secondary School Teachers; Standardized Tests; *Superintendents; *Test Coaching; Testing Problems; Test Results; Test Validity
IDENTIFIERS *Test Directors

ABSTRACT

Two specific questions related to the impact of test preparation materials on test validity were addressed: how extensively test preparation materials are being used; and which materials are being used. Four stratified random samples were drawn from four geographical regions. The 564 subjects in each sample included: (1) school superintendents; (2) testing coordinators; (3) principals; and (4) teachers. A total of 1,262 subjects (56%) responded to a 15-item mail questionnaire referring to norm-referenced tests. Data for 1,012 respondents were analyzed. Sixty-six percent (664) of the respondents used test preparation materials, although teachers reportedly used them less than one would expect based on the reports of superintendents and testing coordinators. Approximately one-third of the 664 respondents who reported using test preparation methods used the "Scoring High" materials, resources that W. Mehrens and J. Kaminski (1989) suggest place the validity of the test in question. The second most often used materials--"Improving Test Taking Skills"--were not linked with a test and should not limit the generalizability of test results. Many educators appeared to use appropriate materials, but concern remains that materials are often used that violate the integrity of test results. It is concluded that in most cases, educators have not jeopardized test valididty. However, the continuing pressure on school districts to perform as well as other districts may cause future problems. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED334249

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

JANIE L. HALL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

PREPARING STUDENTS TO TAKE STANDARDIZED TESTS: HAVE WE GONE TOO FAR?

Janie L. Hall
Oklahoma City, Oklahoma

Paul F. Kleine
University of Oklahoma

Running head: Preparing Students

BEST COPY AVAILABLE

PREPARING STUDENTS TO TAKE STANDARDIZED TESTS:

HAVE WE GONE TOO FAR?

There is concern that increased competition for high scores has encouraged an increase in inappropriate preparation for norm referenced tests. The purpose of this study is to learn the extent to which educators use test preparation materials and to discuss the possible impact of such activities on test validity. We are specifically concerned with 1) teaching test wiseness and 2) teaching to the test (often called teaching the test).

Test Wiseness

Test wiseness, as described by Millman et al (1965, p.707), is the "capacity to utilize the characteristics and formats of the test and/or situation to receive a high score." Sarnacki (1979) described it as a cognitive ability independent of examinees' knowledge of the material being tested. That test wiseness can be taught has been documented by many (e.g., Benson, 1989; Pruettt, 1989; McPhail, 1981).

Properly done, test wiseness is a legitimate part of instruction. Many test and measurement texts include chapters on preparing students for tests (e. g., Thorndike, 1951). And until fairly recently, most teachers took test and measurement courses which included appropriate test preparation (Pechman, 1986). However, a 1984 study found that 60% of the states no longer require test and measurement courses (Wolmut, 1986). From that we draw the conclusion that many educators are not getting advance training in appropriate ways to prepare their students, and are not prepared to distinguish legitimate from illegitimate methods at a

time when they are receiving more pressure to engage in illegitimate activities. (Hall & Kleine, 1990)

Teaching To The Test

Unlike test wiseness, teaching to the test has long had a negative connotation. Although some educators do use the term to describe teaching the domain of information covered by a test, it has more often been considered specific teaching of test content, a practice which can negatively impact generalizability of the results (Mehrens & Kaminski, 1989). Hall & Kleine (1990) found that 92% of respondents to a national survey said their districts have adjusted the district curriculum to the skills measured by the test. Seventy one (71%) said the curriculum had changed, to varying degrees, to match test questions.

Validity

Obviously, the validity of test results becomes an issue in any discussion of test preparation. Tittle (1989) makes an eloquent argument for broadening our concept of validity by pointing out that it is a function of test use and interpretation, as well as construction. We would add test preparation to her definition.

Teaching test wiseness and teaching to the test can have opposite effects on validity (Thorndike, 1951; Mehrens & Kaminski, 1989). Teaching such skills as time use and managing answer sheets can enhance test validity by freeing students to show what they 'really' know. On the other hand, teaching test content negatively impacts the validity of normative comparisons (Shepherd, 1990).

Students coached in the objectives and/or questions on the test are hardly comparable to a group of students who took it 'cold'.

Test Preparation Materials

Mehrens & Kaminski (1989) evaluated several commercial test preparation materials in terms of their match with targeted test content. Using the California Achievement Test (CTB/McGraw-Hill, 1987) for comparison, they found Scoring High on the CAT (Random House, 1987) correlated most highly with the actual test. In their view, the match was so close as to invalidate inference to the broader domain of skills NRTs are designed to sample. Conversely, Riverside's Improving Test Taking Skills (Pritchard, 1983) was the least closely aligned and would not, in the researchers' view, invalidate a broad inference.

Purpose of This Paper

This paper is the second in a series of three based on a national survey of educators. The first (Hall & Kleine, 1990) discussed aligning curicula to match skills and questions on the tests, and test administration practices. The final one will deal with administrative and instructional uses of test results. The current paper addresses two specific questions related to the impact of test preparation on the validity of test results.

1. How extensively are test preparation materials being used?
2. What materials are being used?

Method

Sample

With individuals as the unit of analysis, four stratified random samples (superintendents, testing coordinators, principals, and teachers) were drawn. An independent sampling agency (Market Data Retrieval, 1988), using updated 1980 census data, supplied population figures for each of the four groups. From these, we calculated the sample sizes needed in each cell and asked the agency to draw the samples. Given the size of the populations, a sample size of 376 per group was calculated to provide a 95% confidence level. To increase the likelihood of final groups of sufficient size, we over-sampled by 50%. The final sample size for each group was 564 for a total of 2,256.

Three additional stratifying variables, geographical region, location, and poverty level, were used to ensure that the sample was representative of the nation's educators. The four geographical regions were those defined by the National Education Association (Northeast, Southeast, Midwest, and West). Locations were rural, suburban, or urban. Poverty levels, based on free lunch data, were 1-4%, 5-11%, 12-24%, and 25% or over.

Definitions

A norm-referenced test (NRT) was defined as any nationally standardized, group administered, survey battery designed to sample students' elementary or secondary academic achievement. College admission tests were specifically excluded. Regular teachers were defined as teaching academic subjects in nonspecial classrooms.

Testing coordinators were those individuals whose specific responsibilities include testing activities such as purchase of materials, scheduling, preparation for scoring, and report writing.

We refer to districts as small (under 2,500 students), moderate (2,501 to 9,999 students), moderately large (10,000 to 39,000 students), and large (40,000 or more students). Regions correspond to those defined by the National Education Association. They are; Northeast, Southeast, Midwest, and West.

In discussing appropriateness and inappropriateness of preparation materials, we assumed that the purpose of norm referenced testing is to generalize the results to a broad range of information that students are expected to know at a particular grade level.

Instrument

A group of available administrators and teachers assisted in the initial design of the questionnaire which consisted of 20 items using quantity scales. The instrument was piloted in an Oklahoma public school district using a total of 40 individuals representing the four groups of educators. In addition to answering the questions, respondents described their reactions to the materials and pointed out areas of possible misinterpretation. After revisions based on respondents' comments, a final version consisting of 15 items was administered to all subjects in the study.

The following question addressed use of test preparation materials:

"Which of the following test taking skills materials have you used in the past year? Check as many as apply. (If you are a teacher, answer for yourself. If you are a principal, answer for your school. If you are a superintendent or testing coordinator, answer for your district.)

(1) Locally developed materials.
(2) Scoring High. (American School Publishers, formerly Random House).
(3) Improving Test Taking Skills (Riverside).
(4) A Little Anxiety: Test Taking Skills (CTB/McGraw-Hill).
(5) Other materials (specify).
(6) Do not use test taking skills materials."

We selected Scoring High and Improving Test Taking Skills in order to look at materials clearly identified as appropriate and inappropriate. We also attempted to present materials from each of the other major testing companies, thus the inclusion of A Little Anxiety: Test Taking Skills (CTB/McGraw-Hill). A representative of The Psychological Corporation indicated that at the time of our study his company did not market test preparation materials but they currently do.

In addition to the several groups of educators, we were interested in use of preparation materials as they related to school district size and geographical region. Respondents' region was taken from postal markings on envelopes. An additional question from the survey was included to address size:

"What is the approximate size of your school district?

(1) Under 2,500 students.
(2) 2,501 to 9,999 students.
(3) 10,000 to 39,999 students.
(4) 40,000 or more students.

Procedure

A cover letter specific to each group (e. g., "Dear Superintendent", "Dear Teacher", etc.) requesting cooperation in the study was mailed along with the questionnaire. Both the initial and follow-up questionnaires were mailed first class along with stamped self addressed envelopes. Subjects were informed that the surveys were coded to allow us to follow up on non-respondents. They were assured that their responses would be strictly confidential and that the coding would be destroyed after the second survey was distributed.

To ensure that all respondents interpreted NRTs in the same way, the questionnaire opened with the following statement.

> "The questions below relate to norm referenced tests (NRTs) which, for purposes of this study, are defined as nationally standardized, group administered, surveys of elementary or secondary school subject matter. (College admissions tests are not included.)"

Results

The initial mailing and one follow-up mailing yielded a total return of 1,262 responses, 56% of the total sample. Of these, 221 were superintendents, 421 were testing co-coordinators, 276 were principals and 220 were teachers.

One hundred twenty four (124) respondents, who reported their job as different from the defined groups, were eliminated from the analysis. An additional 94 indicated their district did not give NRTs and 32 did not respond to the question about test preparation materials. These were excluded, leaving a total of 1,012

respondents. Of these, 664 (66%) said they used test preparation materials.

Both superintendents and testing coordinators were asked to respond for their districts. As seen in Table 1, testing coordinators (72%) perceived more use of test preparation materials than superintendents (61%). Teachers (55%) reported less use than the other groups.

Table 1
Use of Test Preparation Materials by Educator Groups

	District		School	Class	
	Supt. n=197	Test. n=392	Prin. n=247	Teach. n=176	Total n=1012*
Use	61%	72%	66%	55%	66%
Don't Use	39%	28%	34%	45%	34%

*Responded to the question.

As will be seen in Table 2, a breakdown by district size (with district, school, and class levels combined) indicates that use of test preparation materials generally increases with size of district from 67% for small to 85% for large.

Table 2
Use of Test Preparation Materials by District Size

	District Size				
	Small n=500	Moderate n=307	Mod. Large n=136	Large n=60	Total n=1003*
Use	58%	69%	77%	85%	66%
Don't Use	42%	31%	23%	15%	34%

* Number who identified district size and answered the question.

When we looked at regions of the country, the Midwest reported more use of test preparation materials (Table 3) than other regions (83%).

Table 3
Use of Test Preparation Materials by Geographical Region

	Region				
	Northeast n=159	Southeast n=297	Midwest n=193	West n=250	Total n=899*
Use	64%	54%	83%	66%	66%
Don't Use	36%	46%	17%	34%	34%

*Number whose region could be identified and who answered the question.

As may be seen in Table 4, Seventy eight percent (78%) of the materials used (by superintendents, testing coordinators, principals, and teachers combined) were locally developed. These were used alone or in combination with other materials. Thirty two percent (32%) use some form of Scoring High materials. Nineteen percent (19%) use the Riverside product. Six percent (6%) use the CTB materials. Eighteen percent (18%) reported using "other" materials. Other could be broken down into four groups; 1) practice tests which accompany the district NRT (17% of the 'Others'), state developed (17%), other commercial materials (23%) and those which could not be defined (43%).

Table 4
Types of Test Preparation Materials Used

	Supt. n=120	Test n=283	Prin. n=164	Teach. n=97	Total n=664*
Local	80%**	76%	83%	72%	78%
Sc. High	22%	42%	29%	19%	31%
Riverside	21%	21%	17%	14%	19%
CTB	9%	4%	3%	1%	6%
Other	8%	22%	15%	24%	18%

* Total reported users.
**Percent of group, e. g., 80% of superintendents used local materials. Columns and rows total more than 100% because respondents used more than one type of material.

The most frequent combination of materials was local and Scoring High (14% of the 664 test preparation material users). Six percent (6%) used local and the Riverside product. Four percent (4%) used local and 'other', and 3% used a combination of local, Scoring High and the Riverside product. Table 5 further illuminates the combinations of test materials that were used. As seen there, the majority of local users used no other materials.

Table 5
Combination of Test Materials Used

	# Users	% Who Use Alone	% Who Use With Other Materials
Local	517	57%*	43%
Sc. High	209	25%	75%
Riverside	125	26%	75%
CTB	29	7%	93%
Other	118	49%	51%

* Percent of users of that type of material; e. g, 57% of the 517 who used local used them alone.

Discussion

Clearly, a majority of our respondents use test preparation materials. Teachers reportedly use them less often than one would expect based on superintendent and testing coordinator responses about districts (We believe the latter represent district policy) However, use of these materials is optional in some districts. It seemed logical that large districts (where one might expect to find more public pressure to perform well) were more inclined to use preparation materials. We can not really explain why the Midwest used more such materials than other regions. We did not look at individual states in deference to respondents who expressed concern about anonymity.

Scoring High is used by approximately one third of all respondents who reported using preparation materials. In light of the findings by Mehrens & Kaminski (1989), these materials are inappropriate when the intent of the test is to generalize to a broad domain of knowledge. Any use of materials such as these places the validity of test results in question and is cause for concern. It is interesting to note that these materials have had several owners in recent years, and even now are owned by a different company that when they were evaluated by Mehrens & Kaminski or when we began our study. There seems to be no change in their format.

Riverside's Improving Test Taking Skills ranked second in usage (19% of all materials users). These materials, as described by Mehrens & Kaminski (1989), are not closely aligned with a test

and should not invalidate broad generalization of test results. Our own examination indicates a high level of instructional utility.

A Little Anxiety, from CTB/McGraw-Hill, is used by 4% of materials users. Our global examination of them indicated to us that they can be considered test wiseness (appropriate) materials.

The 'Other' category of test preparation materials included practice tests which can be purchased along with a specific norm referenced test. These generally provide practice in responding to multiple choice items and seem to be harmless. However, we are not sure how helpful they are. Some educators complain that they are so easy as to give students a false sense of securtiy. State-produced materials seem to fit into the same category of test wiseness as locally developed materials which tend to teach test wiseness and are therefore acceptable.

These findings say to us that a large number of educators are using what we believe are acceptable materials. However, we are still left with the question of how valid test results are when all students do not have the same access to such materials. Additionally, the most widely used commercial materials are products described as inappriopriate. Continued pressure on schools to achieve at high levels suggests the need for close monitoring of use of all these products.

We believe collegial organizations, such as the Association of American Publishers can and should monitor the products of their members and voice disapproval when there are obvious violations of

be more vocal in denouncing materials which violate the integrity of test results.

We continue to argue that educators need to be taught basic measurement information and helped to evaluate the appropriateness of the various materials that are available.

In answer to the question which the title of this paper asks; "...Have we gone too far?", we believe in most cases the answer is no. But we are, at the same time, concerned about the continuing pressure to perform as well as other districts and about the few who do succumb to inappropriate methods. We believe Ligon and Jones were correct when they described an appropriate preparation activity is one which contributes to students' performing on the test near their true achievement levels and is as useful to students as an equal amount of classroom instruction. (Ligon, 1983).

References

Benson, Jeri (1989). The psychometric and cognitive aspects of test-wiseness: a review of the literature. In Michael H. Kean (ed.). Test Wiseness pp. 1-14. Phi Delta Kappa Center for Evaluation, Development, Research, Bloomington, IN.

CTB/McGraw-Hill. (1987). Catalogue of test materials & scoring services. Monterey, CA:Author.

Hall, Janie L. & Kleine, Paul F. (1990) Educator Perceptions of achievement test abuse: a national survey. National Council on Measurement in Education, Boston, MA: April 17, 1990.

Ligon, Glynn D. (1983). Preparing students for standardized tests. In Walter E. Hathaway (ed.) pp. 19-27. New Directions for Testing and Measurement, Jossey-Bass, San Francisco.

Market Data Retrieval (1988). Educational Mailing Lists and Marketing Guide 1988-89. 16 Progress Drive, Shelton, Connecticut 06484.

McPhail, Irving P. (1981). Why teach test wiseness? Journal of Reading, 25(1) October, pp. 32-38.

Mehrens, William A. & Kaminski, John (1989). Methods for improving standardized test scores: fruitful, fruitless, or fraudulent? Educational Measurement Issues and Practice, 8(1) 14-22.

Millman, Jason, Bishop, Carol H., & Ebel, Robert (1965). An analysis of test-wiseness. Educational and Psychological Measurement, 25, pp. 707-726.

Pechman, Ellen (1985). The effects of farming rocky ground under capricious weather conditions. In Peter Wolmut and Grace Iverson (eds.), National Association of Test Directors 1985 Symposia Papers (pp 17-22). Portland, OR:Maultnomah ESD.

Pritchard, F. F. (1983) Improving test-taking skills teacher's guide. (Grades 4-6). Chicago:Riverside.

Pruett, Kathy E. (1989). What does research reveal about the effectiveness of test-taking skills programs? In Michael H. Kean (ed.). Test Wiseness pp. 15-25. Phi Delta Kappa Center for Evaluation, Development, Research, Bloomington, IN.

Random House School Division. (1987). What makes scoring high the most widely used test preparation series (Advertising Brochure 9972620). Westminster, MD: Author.

Sarnacki, Randolph E. (1979). An examination of test wiseness in the cognitive test domain. Review of Educational Research, 4(2) Spring, pp. 252-279.

Shepherd, Lorie A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? Educational Measurement Issues and Practice, 9(3) 15-22.

Thorndike, Robert L. (1951). Reliability. In Everet F Lindquist (ed.). Educational Measurement. Washington, DC, American Council on Education.

Tittle, Carol Kehr (1989). Validity: whose construction is it in the teaching and learning context? Educational Measurement Issues and Practice, 8(1) Spring, pp. 4-13.

Wolmut. Peter (1986). You really oughtta wanna. In Janie Hall and Peter Wolmut (Eds.). National Association of Test Directors 1985 Symposia Papers (pp 18-20). Portland, OR: Multnomah ESD.